Academic Journal of Earth Sciences
ISSN: 2960-0162
DOI: 10.61784/ajesv2n205

APPLICATION OF ENSEMBLE LEARNING IN ADAPTIVE SURFACE MODELING OF SOIL TOTAL POTASSIUM CONTENT IN COMPLEX LANDFORM AREAS

Nikou Heung
Department of Plant, Food, and Environmental Sciences, Faculty of Agriculture, Dalhousie University, Canada.

Abstract: The spatial distribution of soil properties is affected by complex geological environmental factors, and the spatial differentiation characteristics are very obvious. It is difficult to achieve high-precision simulation using a single global interpolation model to simulate soil properties. For the characteristics of spatial discontinuity, limited accuracy of global interpolation models and poor adaptability, this paper proposes an adaptive surface modeling method of soil properties (ASM-SP) supported by ensemble learning and integrating geoscientific environmental variables. Using 110 sample point data collected in 2013, regression kriging (RK), Bayesian kriging (BK), ordinary kriging interpolation (OK), inverse distance weighting (IDW), ASM- SP, the total potassium content of soil was interpolated in Qinghai Lake complex landform type area. This article uses the point-by-point cross validation (LOOCV) interpolation method to simulate accuracy. The results show that ASM-SP not only takes into account the nonlinear relationship between geological environmental variables and soil properties, but also integrates the adaptability advantages of multiple models. It is a new method to achieve high-precision simulation of total soil potassium content in complex landform areas.

Keywords: Spatial interpolation; Adaptive surface modeling; Environmental variables; Linear sweep algorithm; Soil total potassium content; Point-by-point cross-validation

1 OVERVIEW

The total potassium content of soil is the main nutrient element for plant growth, and it is also one of the three elements in the soil that affects crop yield [1]. Studying the spatial distribution characteristics of soil potassium content is not only of great significance for balancing and expanding the soil's available nutrient pool, but also provides a theoretical basis for the sustainable utilization of land resources and the healthy development of regional agriculture [2]. In recent years, foreign Reference [3-4] have conducted in-depth research on the spatial variation patterns of soil; domestic Reference [5] has studied the spatial variation of soil heavy metals and soil organic matter and achieved a large number of research results. The Qinghai Lake Basin is a typical complex landform area. Studying the total potassium content of soil has certain practical significance for improving regional fertilization effects and increasing soil use efficiency [6].

Continuous changes in soil properties are the prerequisite for scientific management and use of soil resources. Spatial interpolation of various soil properties is the main method used to evaluate continuous changes in soil properties [7]. It is also the basis of "digital soil" and "soil metrology" An important research tool. The accuracy of spatial interpolation depends on many factors. Some research results show that the simulation effect of Kriging is better than the inverse distance weighting method (IDW) and Spline, while other research results hold the opposite view [8]. For example, the Reference [9] compared three interpolation methods to evaluate the interpolation error of soil total nitrogen content, and concluded that the radial basis function interpolation model is better than ordinary kriging interpolation (ordinary kriging, OK) and IDW. Xiu Siyu et al. used three interpolation methods to evaluate the annual rainfall of 32 meteorological stations in Heilongjiang Province in 2010. They used average error, average absolute error, average relative error and other indicators to conduct goodness tests, and concluded that the Spline method was optimal and the IDW obtained The maximum and minimum values of rainfall differ the most from the original rainfall data. Reference [10] compared soil Cd pollution interpolation models and concluded that different interpolation methods have different degrees of smoothing for maximum values. Based on the above research findings, current research on spatial attribute surface modeling mainly focuses on a single global interpolation model. Therefore, it is of great significance to study adaptive surface modeling based on spatial attributes.

Reference [11] uses auxiliary variables such as elevation, slope, and land use to predict soil accuracy. Reference [12] used soil conductivity variables to carry out stochastic simulation of spatial variation characteristics of soil salinity at different scales. Reference [13] used NDVI, elevation and distance from the river as auxiliary variables to study salinization in the Yellow River Delta area, and pointed out that the interpolation simulation effect is better when combined with geological environmental variables. However, interpolation models that combine geological environmental variables also have some shortcomings. For example, many studies use a single and similar geological environmental variable to solve prediction problems in different regions. However, different regions are affected by different factors, and the environmental variables considered are different. The adaptability of each interpolation method is different, and there is no absolutely optimal spatial interpolation model. If multi-model integration can be performed in partitions according to the adaptability of different interpolation models, the simulation accuracy can theoretically be improved.

© By the Author(s) 2024, under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0).

1.1 Natural Overview of Qinghai Lake

The study area (36°38'N-37°29'N, 99°52'E-100°50'E) is located in the southeast of the Qinghai Lake basin. Due to crustal movements, complex and diverse landforms have been formed. The landforms are mainly floodplain fan plains. There are more than 40 large and small rivers around the basin. It is an inland closed water system and is a habitat and breeding area for many wild animals. The Qinghai Lake Basin has a semi-arid temperate continental climate on the plateau. It is located in the permafrost zone. It is arid, less rainy, windy, with strong solar radiation and large temperature differences. The study area mainly includes Gonghe, Gangcha and Haiyan. Excluding Qinghai Lake, the total area is about 2100 km2, with an altitude of 3043-4516 m. It has a large number of different soil, vegetation and landform types, and is a typical complex landform. Type area. 1.2 Sample collection

The study conducted soil sampling through spatial layered combination [14]. While sampling the soil, geological environmental information closely related to soil attributes such as the longitude and latitude, altitude, land use type, soil type, vegetation type and geological type of the sampling point was recorded. Each sample point was within 0 to 15 cm, 15 to 30 cm is sampled twice in sequence. In September 2013, with the assistance of personnel from the Qinghai Provincial Environmental Testing Center, the author collected 110 soil surface (0-30 cm) sample point data in typical areas of Qinghai Lake. After sampling, the samples were air-dried, ground, and passed through a 2 mm sieve in the laboratory. The average value of the total potassium content in the soil twice was taken as the recorded sample value.

2 RESEARCH METHODS

2.1 Screening of Geological Environmental Factors

Previous studies have shown that using geological environmental factors as auxiliary variables can effectively improve the interpolation accuracy and mapping effect of soil total potassium content. The driving factors of spatial variability of soil total potassium mainly include: land use type, soil type, vegetation type, geological type, slope, fertilization, etc. [15-16]. Based on the existing research conclusions and combined with this study, which mainly focuses on natural landscape type areas, the two driving factors of fertilization and land management measures were eliminated, and four geological environmental elements of land use type, soil type, vegetation type, and geological type were selected as auxiliary variables. Among them, there are 6 types of land use types including shrubs, cultivated land and grassland; 5 types of soil types include chestnut soil, mobile and aeolian soil; 30 types of vegetation types include water cypress grassland type and Stipa purpurea weed grassland type; There are 13 geological types including sand dunes, sand mountains and valley plains. SPSS was used to calculate the statistical characteristics of soil total potassium content of auxiliary variables. In order to further screen geological elements that have a significant impact on the spatial distribution of soil total potassium, SPSS software was used to perform variance analysis on soil total potassium content and the above four geological elements, and geological elements with significant characteristics were selected as auxiliary environmental variables. The analysis of variance results in Table 1 shows that: in the study area of this article, the landscape fragmentation of vegetation types is too high. There is only one sampling point in many areas, and there are only 1 to 2 sampling points for some subtypes of grassland. The simulation effect is not ideal. Vegetation types are therefore excluded.

2.2 Soil Property Modeling Methods

2.2.1 Selection of interpolation method

The four methods for spatial interpolation of soil potassium content in this article are all implemented based on the ArcGIS 10.2 platform. Traditional interpolation methods include OK, IDW, regression Kriging (RK), and Bayesian Kriging (BK). OK On the basis of satisfying the second-order stationary assumption and intrinsic conditions, and based on the characteristics of the original data and variation function (or covariance) of the regionalized variables, determine the weighted value of the known point parameters around the point to be estimated and the point to be estimated, The advantage of the method of making the optimal estimate of the points to be estimated is that it takes into account the spatial correlation of each sample point, and while obtaining the simulation results, the variance of the estimation accuracy is obtained; IDW uses a linear weight combination of a set of sampling points to determine The output raster value, the weight is inversely proportional to the distance between the interpolation point and the sample point, is a weighted moving average method; RK comprehensively considers a variety of environmental factors that affect the spatial variation of potassium, but due to the distribution of environmental elements in the study area The comparison is fragmented, resulting in not enough sampling points to estimate a relatively accurate semivariance function. Semi-variation analysis was performed on the total potassium content to obtain the parameter values and fitting model during interpolation. The nugget value was small, indicating that the variability caused by its own random error was not large; the S/N+S ratio was close to 1. The N/S ratio is less than 30%, indicating that the total potassium content has strong spatial correlation, and the optimal model for fitting is the exponential model [15].

Table 1 Variance analysis of soil total potassium among different geological element types

Geoscience elements	Source of variance	degrees freedom	ofsum of variances	average variance	F value	P value

	Variance between	106	4．631	0.116		0.038*
land use type	groups	4	0.462	0.044	2．645	
	within group	110	5.093			
	variance					
	total variance					
	Variance between	106	4.371	0.181		
Soil type	groups	4	0.722	0.041	4.378	0．003**
	within group	110	5.093			
	variance					
	total variance					
	Variance between	94	4.159	0.058		
Vegetation Types	groups	16	0.934	0.044	1.319	0.202
	within group	110	5.093			
	variance					
	total variance					
	Variance between	101	4.060	0.115		
geological type	groups	9	1.033	0.04	2．856	0．005**
	within group	110	5.093			
	variance					
	total variance					

Note: * 0.05 significant level; ** 0.01 significant level.

2.2.2 Adaptive surface modeling method integrating geological environment elements

2.2.2.1 Base interpolation model integrating geological environment elements

According to the theory of variation and spatial correlation, the spatial change of any variable can be expressed by the sum of the following two main components: the residual component related to local changes and the structural component related to the trend. The corresponding spatial distribution pattern of soil total potassium content can be expressed as

S(xi,l,k,yj,l,k)= direction(Geox,y)+residual(xi,l,k,yj,l,k) (1)

In the formula, S(xi, l, k, yj, l, k) is the simulated value of the kth type of soil total potassium sampling point of the lth geological element, where (xi, yj) is the sampling point coordinates, i and j represents the row and column of coordinates respectively; direction(Geox, y) is the trend value of the kth type S describing the lth geological feature at (xi, yj), where Geo(x, y) is the trend value describing (xi, yj) geological environment information closely related to soil total potassium at yj); residual(xi, l, k, yj, l, k) is the kth type of the lth geological element describing S at (xi, yj) Residual value of soil total potassium at row i, column j, raster point. The trend function is obtained based on the mean model, and the trend surface describing the structural components is fitted to achieve trend separation; and based on the spatial characteristics of the soil thickness sampling points, the best spatial correlation interpolation algorithm is compared and selected to further process the residuals. Based on the above theoretical knowledge.

2.2.2.2 Using ensemble learning to achieve adaptive partitioning

According to the base interpolation model Modeli established above, the entire study area is simulated to generate a series of soil total potassium interpolation surfaces. The measured values of soil total potassium at the sampling points are subtracted from the predicted values to obtain the simulation error. The sweeping surface line algorithm is used for classification learning of integrated learning. The device scans the interpolation surface, selects scan lines with high classification accuracy and large differences for integration, performs multiple types of adaptive partitioning on each interpolation surface, and obtains the applicable spatial range of each interpolation model.

"+" and "-" respectively indicate sample points that meet the interpolation accuracy requirements and sample points that do not meet the interpolation accuracy requirements. In this process, horizontal and vertical scan lines are used as classifiers to partition the soil total potassium interpolation surface.

(1) Based on the accuracy of the interpolation surface partitioning, a new sample distribution (ie, the weight distribution of each sample point in the sample) D2 and a sub-classifier h1 are obtained. The circled sample indicates that it was misclassified, and the larger "+" indicates that the sample has been weighted.

(2) According to the accuracy of the partition, a new sample distribution D3 and a sub-classifier h2 are obtained. There are 3 "-" symbol classification errors in the weak classifier h2, and the partition error rate ε2 = 0.21 is obtained, and the weight α2 that should be assigned to h2 = 0.66.

(3) Obtain a sub-classifier h3. There are 2 "+" symbols and one "-" symbol in the weak classifier h3, and the partition error rate ε3 = 0.14 is obtained, and the weight h3 should be assigned α3 = 0.91.

(4) Integrate all sub-classifiers to obtain the final Hfinal. After the above three steps, all partitions that meet the accuracy threshold can be extracted. Which category each area belongs to is comprehensively determined by the weight of the classifier where the area belongs. After integration and filling of the NO Data area, the spatial distribution map simulated by adaptive surface modeling for soil properties (ASM-SP) is obtained.

2.3 Accuracy Test of Interpolation Results

The independent verification model has the problem of insufficient testing accuracy, and cross-validation can better evaluate the quality of the interpolation method [17-19]. In the cross-validation, for each of the 110 monitoring points, it is assumed that its data is unknown. Based on the data of the remaining 109 points, different interpolation methods are used to calculate the simulated value, and then the error is calculated based on the observed value.

The evaluation indicators of cross-validation include percentage average estimation error (PAEE), relative mean square error (RMSE), root mean square prediction error (RMSPE), and residual analysis. This article uses maximum error (max error), minimum error (min error), mean error (ME), and root mean square error (RMSE).

3 SIMULATION RESULTS AND ANALYSIS

3.1 Statistical Characteristics of Soil Total Potassium Content

From the analysis of statistical results of potassium content in 110 surface soils in the study area, it can be seen that the value of soil potassium content ranges from 1.406 5% to 2.346 4%, with an average value of 1.958 8%. KS test results It shows that potassium content generally follows a normal distribution. The coefficient of variation of soil potassium content is 10.99%, indicating that the spatial variability of soil potassium content in the study area is at a medium level [15].

3.2 ASM-SP

ASM-SP consists of 3 steps: ①Establishing a series of base interpolation models, namely OK-Landuse, OK-Soil, OK-Geology; ② Adaptive segmentation of base interpolation surfaces; ③ Optimal combination of base interpolation. For details on the adaptive partitioning method, see Section 2.2. The ASM-SP building process is as follows:

3.2.1 Establishment of base interpolation model

(1) Calculate the average soil total potassium value for each geological element and obtain the trend surface direction (Geox, y). Based on the measured soil total potassium value and combined with auxiliary variables, the soil total potassium average value is obtained.

(2) Subtract the average soil total potassium content from the measured value to calculate the simulated residual of soil total potassium. Perform OK interpolation on the simulation residuals to obtain the residual surface residual(xi, l, k, yj, l, k).

(3) Use formula (1) to calculate the simulated value S (xi, l, k, yj, l, k) of the k-th type of soil total potassium sampling point of the l-th geological element, that is, the above-mentioned trend surface and residual Surfaces are added. This is the construction process of the base interpolation surface.

3.2.2 Adaptive surface segmentation based on linear scanning algorithm

Based on the method of constructing error surfaces described in Section 2.2.2, the error surfaces of different interpolation models are obtained, and the applicable range of each interpolation model is determined.

3.2.3 Integration of interpolation surfaces

On the basis of grid unit optimization, the interpolation result of the grid unit with the smallest error is selected as the best grid unit for integration. Optimal partitioning corresponding to different interpolation models is shown.

3.3 Analysis of Prediction Accuracy of Different Methods

In order to evaluate the accuracy of ASM-SP in simulating the spatial distribution of soil total potassium content, this paper compared the accuracy of five interpolation methods, namely RK, BK, IDW, OK+Landuse and ASM-SP. OK and BK are combined with the auxiliary variable Landuse to compare with ASM-SP; the RK method is combined with Landuse to fit the trend term. The error index of point-by-point cross-validation (see Table 2), for the maximum error value, the small value is high and the precision is high, and for the minimum error value, the small value is high precision. It shows that ASM-SP has the smallest prediction error range, and the sensitivity of prediction values and the ability to reflect extreme values are better than the other four methods. For the values of ME and RMSE, ASM-SP reaches -0.002 1 and 0.281 1 respectively, which is better than other results, and the accuracy AC value is the best, so ASM-SP shows better performance , its ME value is closer to 0 than that of traditional interpolation methods (ie OK and IDW). This means that interpolation combined with auxiliary variables is more unbiased. And the accuracy of ASM -SP also fully shows that its regression curve can better simulate the relationship between predicted values and true values. In summary, the interpolation method of ASM-SP is the best.

In general, the main reasons why the ASM-SP interpolation method is better than other methods are: ① This method combines auxiliary variables, so it more accurately depicts the boundary of soil total potassium that changes with geological environmental factors; ② Uses a linear scan algorithm By dividing the optimal area and integrating each geological element into the optimal area according to this method, the accuracy of interpolation is improved to a great extent.

Table 2 Cross-validation results of interpolation effectiveness of 5 methods

interpolation method	max error	min error	ME	RMSE	AC
RK	5.141 2+E01	2.750 8-E02	0.007 3	0.321 3	0.927 3
BK	5.189 9+E01	2.345 5-E02	- 0.005 3	0.219 7	0.918 7

IDW	5.549 1+E01	3.450 7-E02	0.009 2	1.056 7	0.875 9
OK+Landuse	5.231 1+E01	1.231 9-E02	0.008 7	0.978 8	0.897 7
ASM-SP	5.123 1+E01	2.231 1-E02	- 0.002 1	0.281 1	0.931 5

3.4 Comparison of the Effects of Different Prediction Methods

This paper compares the interpolation effects of total potassium content of five types of soil (shown in Figure 7). It can be seen from Figure 7 that the degree of variation depicted by RK, BK and OK is smaller than the true value interval, and the interpolation shows varying degrees of weak "bull's eye" effects. IDW interpolation can reflect the overall pattern of soil total potassium distribution, but it has a strong "bull's eye" effect and low interpolation accuracy.

ASM-SP can effectively characterize the spatial variation pattern of soil total potassium, generate a moderate interpolation range (1.30-2.32), and can reflect the details of local changes with more spatial differentiation of soil total potassium. The soil total potassium value obtained by BK, RK and OK interpolation and IDW interpolation cannot describe the mutation boundary information caused by changes in soil properties with geological environmental variables. The ASM-SP method has strong adaptability to the spatial interpolation of soil properties in areas with complex topography and can describe it more accurately.

4 CONCLUSION

Using geological environmental variables to correct the interpolation results can solve the problem of spatial differentiation and mutation of soil attributes, especially the problem of large mutations in complex landform areas such as hills and gullies and in areas where different geological elements are connected. Spatial differentiation of soil properties produces large changes over short horizontal distances, resulting in limited accuracy of a single global interpolation model. Therefore, this paper proposes an adaptive modeling (ASM-SP) method that integrates geological environment elements and performs adaptive modeling of the optimal areas of different interpolation models. In areas with complex landform types, ASM-SP can more accurately characterize the spatial variation of soil properties and effectively reduce simulation errors. In addition, ASM-SP combines auxiliary variables to make the simulation results more consistent with geological laws and facilitate physical explanation of the spatial variation characteristics of soil properties. Comparing the results of spatial interpolation models (IDW, OK) that do not use auxiliary variables, and the RK method that uses topographic factors to assist interpolation, it is shown that ASM-SP depicts soil total potassium content more in line with the spatial variation patterns of soil attributes in the study area, and the edge of geological elements Detailed information is more obvious. Accuracy evaluation indicators such as ME and RMSE are also smaller, reaching -0.002 1 and 0.281 1 respectively, showing higher simulation accuracy than other interpolation models, and can especially accurately depict the spatial variation of soil total potassium in complex landform types. The changing and prominent boundaries of the surrounding geological environmental elements. ASM-SP is a new method to achieve high-precision simulation of soil total potassium content in complex landform areas.

The research on soil attribute mapping provides new ideas and useful reference.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Chen Ge, Tang Chunchun, Li Zusheng. Effects of different fertilization measures on dryland fertility and crop yield in Dongting Lake Plain area. Chinese Journal of Ecological Agriculture, 2017, 25(5): 689-697.

[2] Dong Hongfang, Yu Junbao, Sun Zhigao. Spatial distribution characteristics of organic carbon in the plant-soil system of tidal flat wetlands along the Yellow River estuary. Environmental Science, 2010, 31(6): 1594-1599.

[3] OBALUM SE, OPPONG J, IGWE CA. Spatial variability of uncultivated soils in derived savanna. International Agrophysics, 2013, 27(1): 57-67.

[4] ROSEMARY F, VITHARANA UWA, INDRARATNE SP. Exploring the spatial variability of soil properties in an Alfisol soil catena． Catena, 2017, 150: 53-61.

[5] Zhao Mingsong, Zhang Ganlin, Wang Decai. Analysis of spatial variation characteristics and main controlling factors of soil organic matter in the Xuhuai Yellow Flood Plain. Journal of Soil Science, 2013, 50(1): 1-11.

[6] Long Jun, Zhang Liming, Shen Jinquan. Research on spatial interpolation method of soil organic matter in cultivated land in complex landform areas. Acta Soil Sinica, 2014, 51 (6): 1270-1281.

[7] Shi Wenjiao, Yue Tianxiang, Shi Xiaoli. Research progress on spatial interpolation methods and accuracy of continuous soil properties. Journal of Natural Resources, 2012, 27(1): 163-175.

[8] Ma Chengxia, Ding Jianli, Wang Lu. Research on interpolation method for spatial variation analysis of oasis soil surface salt content. Soil and Water Conservation Research, 2014, 21(4): 317-320.

[9] Zhao Qiaoli, Zheng Guoqing, Feng Xiao. Comparative analysis of three spatial interpolation methods of soil total nitrogen content in Anyang County, Henan Province. Soil Bulletin, 2012, 43(5): 1162-1166.

[10] Xie Yunfeng, Chen Tongbin, Lei Mei. The influence of spatial interpolation model on soil Cd pollution assessment results. Journal of Environmental Science, 2010, 30(4): 847-854.

[11] KURIAKOSE SL, DEVKOTA S, ROSSITER DG. Prediction of soil depth using environmental variables in an anthropogenic landscape: a case study in the Western Ghats of Kerala, India. Catena, 2009, 79(1): 27-38.

[12] Jiang Guirong. Spatial variation characteristics and uncertainty analysis of soil salinity at different scales in arid areas. Beijing: China University of Geosciences, 2012.

[13] Wu Chunsheng, Huang Chong, Liu Gaohuan. Research on spatial prediction method of soil salinity in the Yellow River Delta. Resource Science, 2016, 38 (4): 704-713.

[14] Yang Lin, Zhu Axing, Zhang Shujie. Comparative study of multi-level representative sampling and stratified random sampling in soil mapping, Acta Soil Sinica, 2015, 52 (1): 28-37.

[15] Wang Shengli, Liu Wei, Zhang Lianpeng. Adaptive surface modeling of soil total potassium content supported by geological environmental variables - taking the typical area of Qinghai Lake Basin as an example. Soil and Water Conservation Research, 2018, 25(1): 132-138.

[16] Huang Wenzhong. Study on the spatial variation characteristics and influencing factors of soil potassium in Yibin City. Ya'an: Sichuan Agricultural University, 2010.

[17] Xu Aiping, Sheng Wenshun, Shu Hong. Data interpolation and cross-validation of space-time product sum model. Journal of Wuhan University (Information Science Edition), 2012, 37 (7): 766-769.

[18] Li Jia, Duan Ping, Lu Haiyang. RBF morphological parameter optimization method based on improved point-by-point cross-validation and its spatial interpolation experiment. Geography and Geographical Information Science, 2016, 32(3): 39-42.

[19] Gu Chunlei, Yang Yang, Zhu Zhichun. Cross-validation of the accuracy of several interpolation methods for establishing DEM models. Surveying, Mapping and Spatial Geographic Information, 2011, 34 (5): 99-102.